Getting Started with R for Mass Spectrometrists

ASMS - Short Course

Jeff Jones, Ph. D.
Heath Patterson, Ph. D.

Main Topics for the Short Course

Learn …

- basic fundamentals of the **R programming language**

- how to use to the **R Studio** integrated development environment (IDE)

- about **tidy data**, what it is, and why it's important for data analysis

- fundamentals of the **tidyverse** ecosystem of R packages and how they can be used to streamline the data analysis process

- how to make data visualizations using the **ggplot2** R package

- about some fundamental R packages for **mass spectrometry**

The Team

Jeff Jones Ph.D.

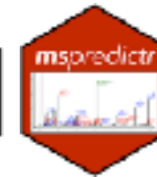Senior Scientist Proteomics, Div. Physics, Math and Astronomy, Caltech

"Started programming out of a need to analyze large data sets generated by mass spectrometry."

in linkedin.com/in/jeffsocal/   github.com/jeffsocal

tidyproteomics jeffsocal.github.io/tidyproteomics
An R package for the tidy-ing, post processing and analysis of quantitative proteomic data.

Other
Mass Spec.
R Packages

Nathan "Heath" Patterson Ph.D.

Director of Spatial Biology Bioinformatics, Aspect Analytics, Genk, Belgium

"Took an interest in programming when vendor solutions no longer solved the problem"

in linkedin.com/in/heath-patterson-a2402a100/   github.com/nhpatterson

---

Ryan Benz Ph.D.

Director Data Science, Seer Inc.

"Learning to code, even at a basic level, will open up a huge amount of new possibilities for any scientist"

in linkedin.com/in/ryan-benz/   github.com/ZenBrayn

Why Learn R?

You choose wisely ... R is a great language for data analysis!

- Many programming languages are general purpose (can be used in any domain), e.g. C/C++, Java, Python

- R is not a general purpose programming language, it's a language specifically designed for working with data (that's what scientists do!)

- Because R is geared toward data, its design, structure and continued development is focused on making it easier to work with data

- R has become one of the top languages for data science, and it's popularity and usage continues to grow

- For scientists, R is a great tool to learn

Write Code with *R* and *RStudio IDE*

- RStudio provides an integrated development environment (IDE) for coding.
- Streamlines all aspects of the data analysis process.
- Learn best practices for organizing analyses into projects and make your work reproducible.
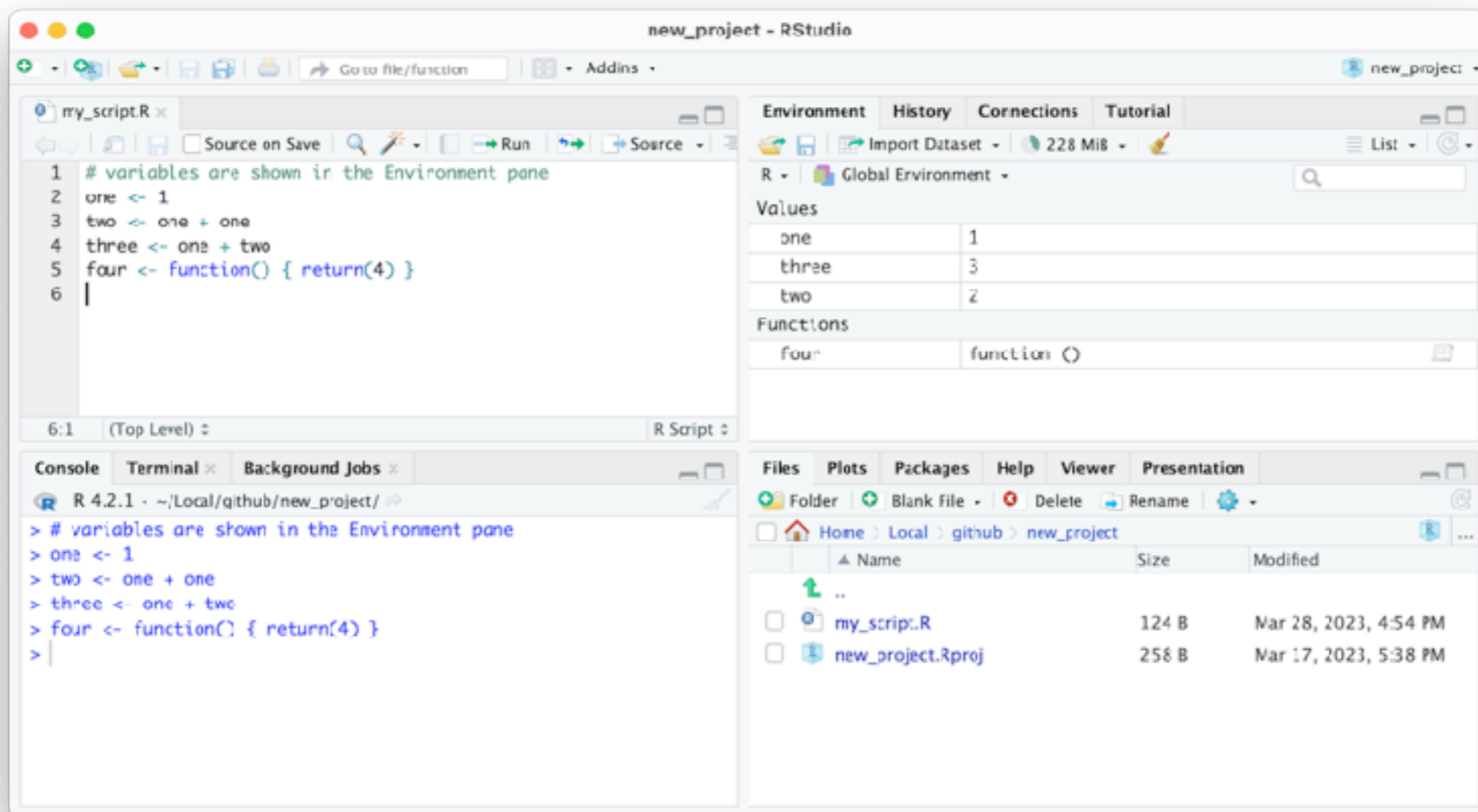
Organize Data with the *Tidyverse*

- Learn how *clean-up* data and make analysis code easier to read, write and understand.
- The tidyverse covers all steps of the analysis process, from data import & cleaning to visualization and final reporting
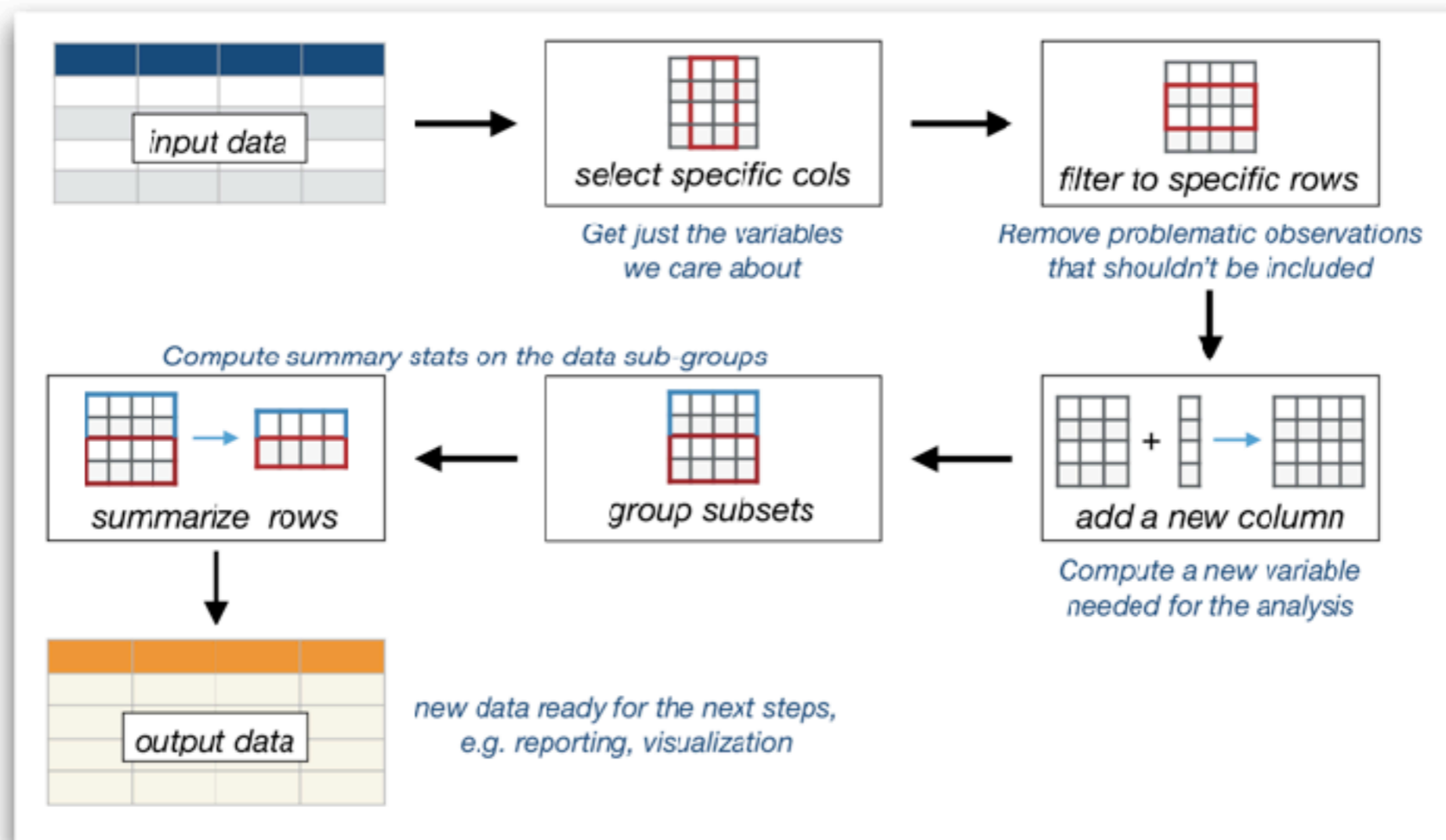
Visualize Data with *GGPlots*

- Lean to visualize data helping to understand patterns, trends, and relationships.
- A well-designed visualization can make complex data easy to understand and convey insights that would be hard to discern from raw data.
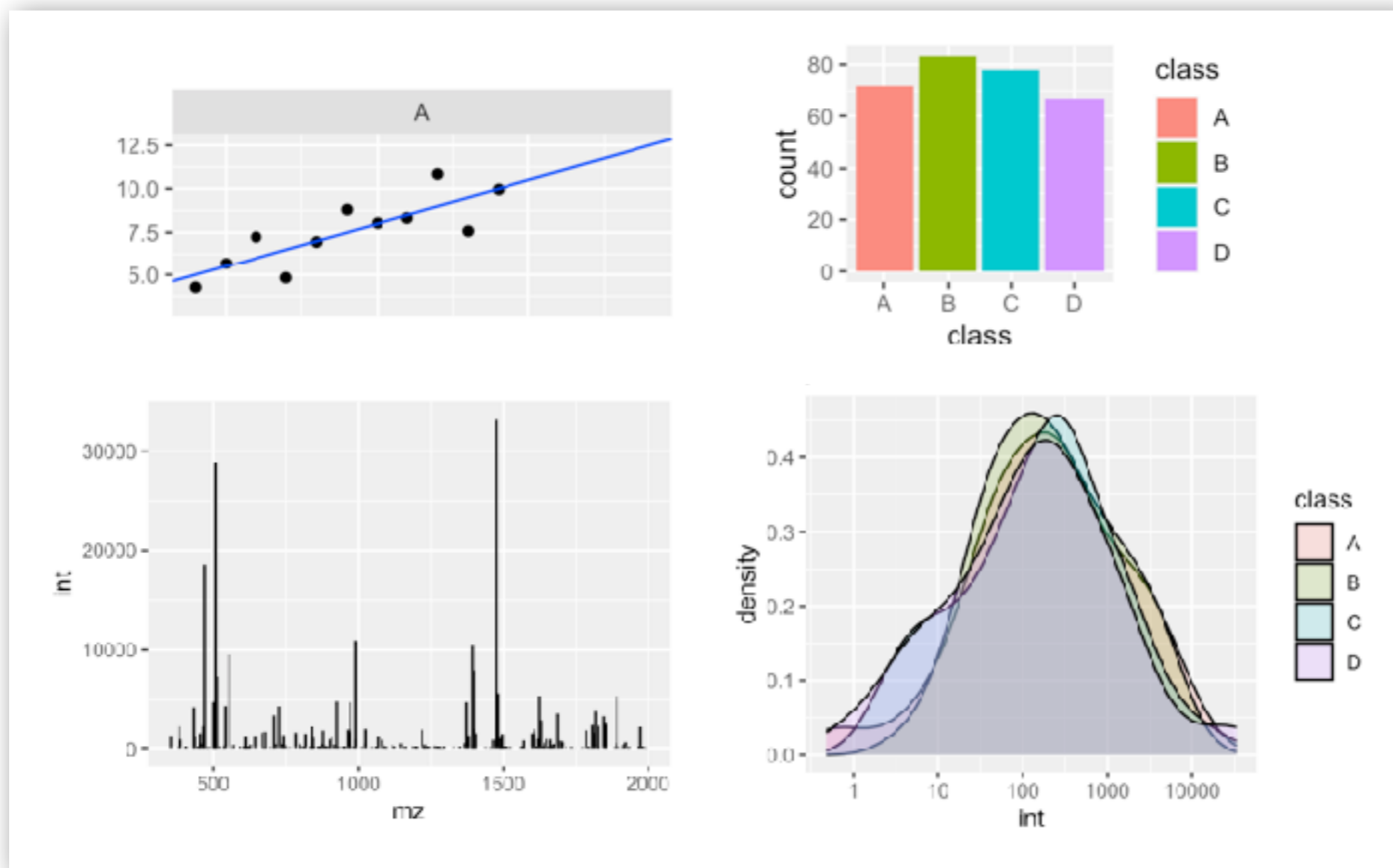
What do we hope to accomplish?

At the end of the course, you are able to …

- start-up R Studio and make an R project.

- read a formatted data file in to R.

- understand basic properties about the data.

- convey what tidy data is and why it is important.

- perform data manipulations, summaries and analyses.

- construct a plot with the data.

- know where to go to expand your knowledge.